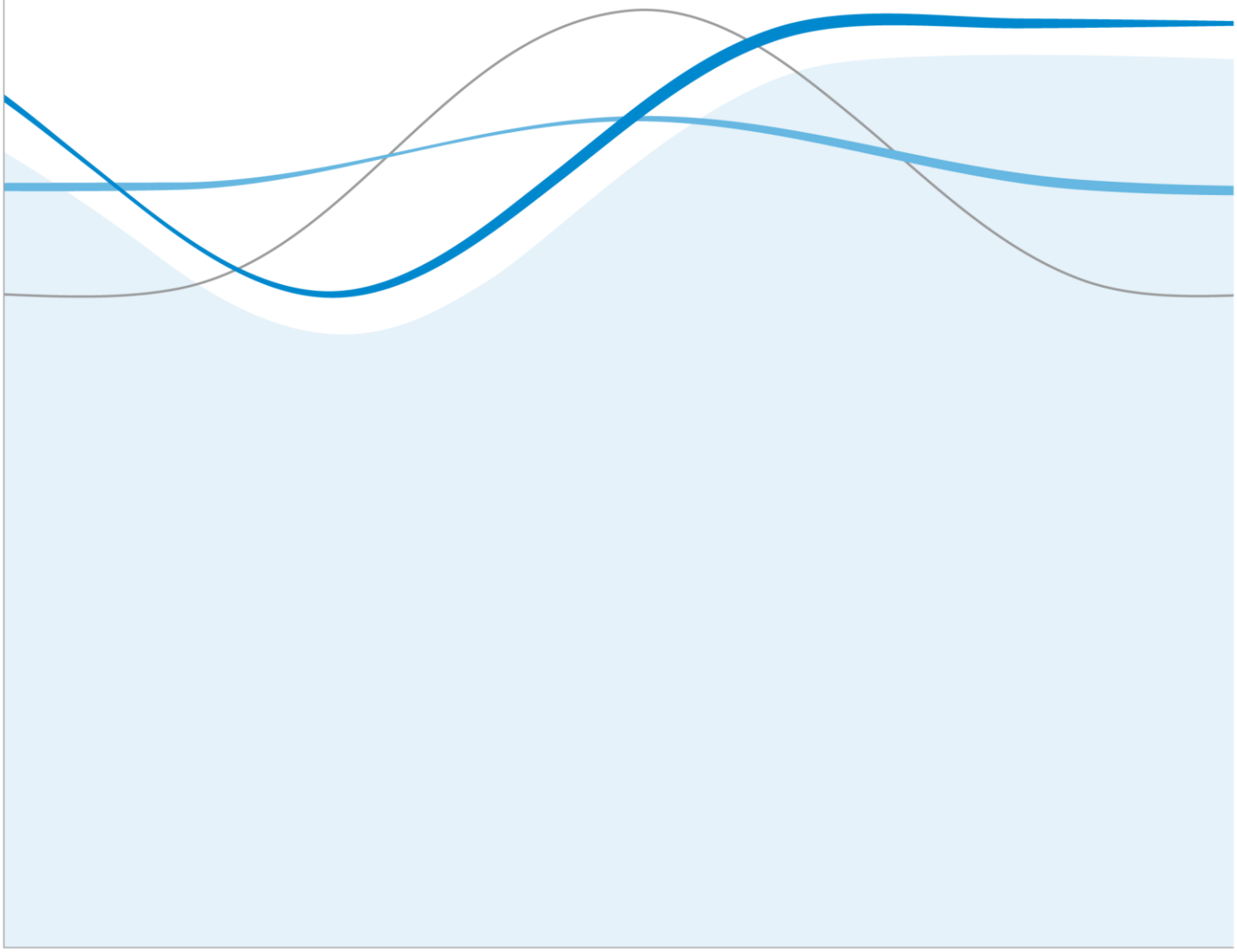
Office of Health Economics Consulting

Consulting Report

Shaping the Research Agenda to Estimate Relevant Cost-effectiveness Thresholds for Health Technology Assessment Decision Making: Report for ABPI

April 2016

Sarah Karlsberg Schaffer, Patricia Cubi-Molla, Nancy Devlin and Adrian Towse



Shaping the Research Agenda to Estimate Relevant Cost-effectiveness Thresholds for Health Technology Assessment Decision Making: Report for ABPI

Sarah Karlsberg Schaffer, Patricia Cubi-Molla, Nancy Devlin and Adrian Towse

April 2016

Submitted by:
OHE Consulting Ltd
(a registered company number 09853113)
Southside, 7th Floor
105 Victoria Street
London SW1E 6QT
United Kingdom

For further information please contact
Sarah Karlsberg Schaffer
Economist
Tel: +44(0)207 747 8862
Email: sschaffer@ohe.org

About OHE Consulting Reports

Many of the studies OHE Consulting performs are proprietary and the results are not released publicly. Studies of interest to a wide audience, however, may be made available, in whole or in part, with the client's permission. They may be published by OHE alone, jointly with the client, or externally in scholarly publications.

Studies published by OHE as OHE Consulting Reports are subject to internal quality assurance but do not go through the OHE Editorial Board peer review process. Publication is at the client's discretion.

Table of Contents

Key messages

- Interviews and a day-long workshop were conducted with a selection of experts in the fields of health and welfare economics in order to discuss the UK cost-effectiveness threshold; examine how the threshold is currently conceptualised, estimated and used in HTA decision making; and discuss how measures of value other than quality-adjusted life years (QALYs) could be incorporated into the decision making framework.
- The primary objective of the workshop was to begin to shape a research agenda around future estimation and use of the threshold and to explore how other types of value could be incorporated into the UK HTA process.
- All participants shared the view that the topic is extremely important – and that it is very desirable to improve the evidence base for decisions regarding resource allocation and patients' access to health care.
- There was a consensus view that there is no single best or correct approach to empirically estimating the threshold – and that therefore generating complementary evidence from a range of methods would be an appropriate way of informing the selection of cost effectiveness thresholds for use in HTA decisions.
- There was debate among participants over whether researchers should aim to identify just one threshold, as opposed to a threshold range within which there is opportunity for deliberation (similar to the current method used by NICE).
- Local prioritisation frameworks could be developed in order to strengthen commissioning and aid resource allocation decisions, and to ensure greater consistency across local decision making, and between local and national decision making.
- Disinvestment was identified as a key challenge to local NHS decision makers, partly because there is little evidence on the cost per QALY of non-pharmaceutical interventions.
- It was generally agreed that the HTA process would benefit from a more transparent approach to decision making than is used currently. However there was no clear consensus regarding the use of MCDA as part of decision making, with some individuals advocating its use in HTA and others expressing their perception that MCDA may be too algorithmic, and would reduce the role of deliberation.
- Given the information gathered in the interviews and the workshop, OHE recommends the following six potential directions for future research:
    - Inferring the threshold from the output of PBMA (programme budgeting and marginal analysis) exercises;
        - Potential to understand what is being traded-off by decision makers and the criteria used to make these decisions;
    - Examining the clinical threshold;
        - Potential to estimate the threshold by identifying, for a particular treatment or procedure, the point at which a clinician makes the decision to treat or not treat a patient with an intervention;
    - Exploring differences in thresholds across clinical areas;

        - Evidence suggests that the cost per QALY gained varies hugely across programme budgeting categories; understanding this variation could help to inform HTA decisions at the national level;
    - Further exploration of econometric analysis of NHS data;
        - Specific approaches include the use of panel data (expenditure and health outcome data for the same health organisations over multiple time periods);
    - Improving evidence on social values;
        - Expanding the evidence base on social values relevant to health care. This might include disease severity and rarity, socioeconomic equity, and other patient population characteristics;
    - Structured decision making in practice;
        - Potential to test the use of existing techniques for structured decision making, for example in mock technology appraisals for hypothetical new products (using current/former Committee members).

The six key research areas identified could help shape the research agenda to optimise reform of the UK HTA decision making framework, and should be prioritised as research topics to further inform this important debate.

Background and objectives

The level at which the cost-effectiveness threshold is set is of huge importance because it affects the decisions of NHS bodies, most notably NICE, regarding the adoption of new health care technologies. For cost increasing technologies, this affects resource allocation in health care, giving rise to opportunity costs of the health foregone from other uses of those resources. In current financial circumstances, with very small real increases in health budgets, this trade-off has become even more acute.

There have been a number of studies aimed at estimating the threshold including, most notably, Claxton and colleagues' 2015 *Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold*, published in *Health Technology Assessment,* which estimated the threshold to be around £13,000 per QALY in 2008[1,2], substantially lower than the threshold range currently used by NICE (£20,000-£30,000 per QALY).

Given the significance of the threshold for UK health and pharmaceutical industry policy, and the level of attention given to the estimates produced by Claxton et al., ABPI commissioned OHE to explore the academic landscape for views on how research on the cost-effectiveness threshold can be developed.

The objectives of this project are to:

---

[1]Claxton K, Martin S, Soares M, Rice N, Spackman S, Hinde S, Devlin N, Smith P, Sculpher M. (2015). Methods for the estimation of the National Institute for Health and Care Excellence NICE cost-effectiveness threshold. CHE Research Paper 81. Revised report following referees comments. York: University of York. *Health Technology Assessment.* 19;14.

[2] A previous version of this paper was the subject of an OHE critique: Barnsley, P., Towse, A., Karlsberg Schaffer, S. and Sussex J. (2013). *Critique of CHE Research Paper 81 "Methods for the estimation of the NICE cost-effectiveness threshold".* OHE Occasional Paper 13/07. London: Office of Health Economics.

- Advance the discussion on the UK cost-effectiveness threshold; to examine how the threshold is currently conceptualised, estimated and used in HTA decision making; and to discuss how measures of value other than quality-adjusted life years (QALYs) could be incorporated into the decision making framework.
- Collect the views of leading experts in the academic and policy community in the UK about potential new directions for additional research, and to develop a report summarising the "lay of the land" for potential research in this area;
- Begin to shape a research agenda for contributing to the academic and policy debate in future, potentially in collaboration with the economists involved in this project if these are topics of research interest to them.

In order to achieve these objectives, OHE conducted interviews with a selection of experts in the field. OHE then analysed the transcripts of these interviews to produce summary statistics showing the proportion of participants who agreed with a number of dichotomous statements regarding the research topic. The second key output of the interviews was a list of research directions highlighted by the interviewees.

Following the interviews, OHE held a workshop with a number of the experts during which participants discussed the output of the interviews and prioritised possible future research that could contribute to developing methodologies and/or empirical evidence to drive forward the policy debate.

The interviews

Selection of interviewees

In early 2015, OHE created a shortlist of universities from which individual academics would be contacted for invitation to participate in the project, using the following criteria:

- An active academic department with a significant number of economists publishing in the field of health economics, particularly the economics of HTA;
- A focussed and energetic health economics unit or group;
- The resources, reputation and interests to challenge orthodox approaches to health economics;
- Staff with connections across the health care and health policy sector, not limited to health economics or academia.

From each university, a selection of academics was chosen based on their expertise in health economics and related areas such as welfare and behavioural economics. Given that the views of the authors of *Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold*[3] are already well known in the academic community, we specifically sought alternative and novel ideas for research to inform selection of the cost-effectiveness threshold to use in HTA.

OHE invited 23 individual academics to participate in the project. Table 1 reports the names and affiliations of the 15 who took part, as well as an indicator to show which of the interviewees also attended the workshop. Of the remaining eight, three indicated their willingness to participate but were unavailable, three did not reply to invitations and two declined.

[3] http://www.journalslibrary.nihr.ac.uk/hta/volume-19/issue-14#abstract

Table 1. Project participants

Name	Affiliation	Attended workshop?
Alastair Gray	University of Oxford	Yes
Alistair McGuire	London School of Economics	No
Ben van Hout	University of Sheffield	Yes
Carol Propper	Imperial College London	No
David Parkin	King's College London	No
Dyfrig Hughes	Bangor University	No
Graham Loomes	University of Warwick	Yes
Joanne Lord	University of Southampton	Yes
John Appleby	The King's Fund	No
John Cairns	London School of Hygiene & Tropical Medicine	Yes
Jose Luis Pinto Prades	Glasgow Caledonian University	Yes
Martin Buxton	Brunel University London	Yes
Mireia Jofre-Bonet	City University London	Yes
Richard Cookson	University of York	Yes
Stephen Birch	McMaster University	No

Interview methods

Semi-structured interviews were completed by Sarah Karlsberg Schaffer and Patricia Cubi-Molla using the agreed interview guide.[4] The key sections of the interview guide are summarised below:

1. Interviewee's past and future research on the cost-effectiveness threshold
2. Conceptualising the threshold
    - In your view, what does the cost-effectiveness threshold to be used in HTA decisions represent?
3. Estimating the threshold
    - What are the challenges in estimating what the threshold should be?
    - In a perfect world, with perfect data, how would you design a study with the aim of estimating the true cost-effectiveness threshold in the UK?
    - What is your understanding of how NHS funding decisions are made in practice?
4. Alternative methods of HTA and priority setting in health care
    - Should measures of value other than, or addition to, the QALY (such as the social value involved in addressing a severe or rare condition) be accommodated in HTA processes?
    - How can the threshold reflect societal preferences such as those for treating patients with severe diseases?
    - Should HTA involve assessing the cost-effectiveness of existing services?
    - Is there a role for multi-criteria decision making analysis (MCDA) to consider measures of value beyond the QALY in HTA?
    - Is there a role for formal priority setting frameworks e.g., Programme Budgeting and Marginal Analysis (PBMA) in local NHS resource allocation decisions?

[4] PCM was present for all but two interviews.

5. Any additional points the interviewee wishes to make about the theory or empirical approach to the selection of the cost-effectiveness threshold.

Interviews lasted between 40 and 90 minutes; two were performed face to face and 13 by telephone. All interviews were recorded and transcribed to ensure efficient and accurate capture of information. In most cases, questions were "open" in that a list of potential answers was not given (some questions in Section 4 of the interview guide shown above have yes/no answers).

Consensus of interviewees regarding dichotomous statements

Informed by the transcripts and structured according to the interview guide, a number of dichotomous statements were identified by SKS from the interviews, and reviewed and refined by PCM (see Figure 1). Each transcript was then coded using qualitative research software, ATLAS.ti (version 7) – eight transcripts were coded by SKS and seven by PCM. For each of the statements, codes were created for positions "in favour" and "against". Codes were also created for quotes to illustrate important themes. In cases where interpretation of a transcript was not clear, SKS and PCM reviewed the statements and came to a consensus.

Figure 1 displays a selection of interview results, specifically the proportion of interviewees agreeing with particular statements. Note that the percentages do not sum to 100 as some interviewees did not express an opinion on some statements. Figures in blue are percentages conditional on the answer to the previous question, and key quotes are contained in boxes throughout the following pages.

For example, 93% of interviewees were of the view that HTA should go beyond the ICER as the only decision criterion. Participants were not asked their views on particular decision criteria, although the interviewers gave the examples of severity and rarity of disease (see question 4 of the interview script, above). Therefore, only a selection of interviewees named each of the criteria listed in Figure 1, and the percentages should be interpreted as conservative. Of the 93% who believed that other criteria should be involved, 38% believed equity should be included, 15% believed equity should not be included and the remainder (47%) expressed no view. 15% of participants believed that the innovation of the new technology should be considered as part of the HTA process, with no participants expressing the opposite view. 23% were in favour of considering the severity of disease, whereas 15% were against this.

The results suggest that the majority of interviewees believe that the cost-effectiveness threshold should be *conceptualised* as the shadow price of the proposed investment. In other words, the majority expressed that it represents the opportunity cost given a fixed health care budget.

> *"[The threshold] should be interpreted [as] the shadow price of the budget constraint, which just means how many QALYs, how much health, you lose when you take money out of the NHS to pay for your new drug or your new intervention."*

A large majority believed that the threshold should be *estimated* using the shadow price approach, as compared to the willingness-to-pay (WTP) approach. Note that although a sizeable proportion believed that the WTP conceptualisation of the threshold was "correct", a number of these interviewees believed that the shadow price approach was more achievable in terms of estimating it.

"Particularly in an environment where the health budget is, basically, not growing in real terms … Any new treatment which is being made available can only really happen either by making efficiency savings or by stopping doing other things which are less cost-effective"

Figure 1. Consensus of interviewees regarding dichotomous statements

The threshold should represent…	
The willingness-to-pay (WTP) of society	21%
The shadow price opportunity cost of the investment	64%
The threshold should be estimated using…	
The shadow price approach	86%
The WTP approach	7%

	In favour	Against
HTA should go beyond the ICER as the only decision criterion	**93%**	7%
"Equity" included among the decision criteria	38%	15%
"Innovation" included among the decision criteria	15%	0%
"Severity" included among the decision criteria	23%	15%
"Productivity loss" included among the decision criteria	8%	0%
"QoL of carers" included among the decision criteria	8%	0%
"Uncertainty" included among the decision criteria	8%	8%
"Rarity" included among the decision criteria	8%	23%
"End of Life" included among the decision criteria	0%	15%
"Cancer" included among the decision criteria	0%	23%

HTA should reflect social preferences using...	**In favour**	**Against**
A deliberative process	36%	14%
Weighting QALYs or thresholds (e.g. for severity or equity)	14%	43%
Applying weights to different criteria	**43%**	29%
Weighted sum	0%	17%
Programme budgeting and marginal analysis (PBMA)	67%	17%
Portfolio Analysis	17%	0%

	In favour	Against
HTA should involve assessment of cost-effectiveness of existing services	50%	7%

Almost all participants believed that HTA should go beyond the ICER as the only decision criterion.[5] For some participants (to varying degrees), this reflected the belief that the QALY is an imperfect measure of health gain. For others, it related to the notion that

[5] See "Dakin, H., Devlin, N., Feng, Y., Rice, N., O'Neill, P., & Parkin, D. (2015). The Influence of Cost-Effectiveness and Other Factors on NICE Decisions. *Health economics*, *24*(10), 1256-1271" for a discussion of the factors that have affected NICE committee decisions, and the challenges associated with identifying these factors and their relative influence.

even if the QALY *was* a perfect measure of health gain, providing QALYs to certain individuals (e.g. those at the end of their lives) is inherently likely to be viewed as more valuable by society. The decision making criterion favoured by most interviewees was "equity", generally considered in socioeconomic terms. The criteria least favoured by interviewees were rarity of the disease and treatments for cancer specifically (generally referring to the Cancer Drugs Fund).

> *"There is a need to recognise that QALYs are limited in their capacity to reflect health or improvement in the NHS"*

Of those that believed that social values beyond the QALY should be reflected in HTA, around a third believed this should be accommodated as part of a deliberative process (similar to how NICE Appraisal Committees operate currently, apart from the use of the end-of-life criteria).

> *"I think that quantifying things to the n^{th} degree just adds complications, confusions and impacts on transparency"*

About 40% of interviewees believed that HTA should use explicit weighting of different criteria. This refers to some form of multi-criteria decision analysis (MCDA), where "QALYs produced" is one attribute of value that can be weighted and combined with others (for example, severity or rarity of disease) and compared against some threshold. Around a third (36%) of participants believed that consideration of criteria beyond the ICER should remain a deliberative process.

The various MCDA methodologies discussed by interviewees were a simple weighted sum, programme budgeting and marginal analysis (PBMA) and portfolio optimisation (an extension of PBMA using mathematical programming models[6]). This MCDA type approach is contrasted with QALY weighting, whereby the QALYs used to calculate the ICER are weighted according to social values. This is equivalent to the cost-per-QALY threshold varying according to these values (similar to the way it does now for end-of-life diseases). In the UK value-based pricing discussion, a disease severity weighting for the QALY was proposed by the DH.

> *"We want the [HTA] process to be as explicit as possible and I think MCDA done very well may help to make some of the trade-offs explicit but done badly, it becomes an even worse form of box filling exercise"*

Finally, 50% of interviewees expressed the view that HTA should involve the assessment of the cost-effectiveness of existing services, as opposed to only new technologies. Those who were in favour of doing so focussed on the imbalance in the treatment of new medicines versus all other technologies/services provided by the health service. In addition, it was mentioned that one can only identify the opportunity cost of NICE

[6] Birch, S., & Gafni, A. (2015). On the margins of health economics: a response to 'resolving NICE'S nasty dilemma'. *Health Economics, Policy and Law*, *10*(02), 183-193.

recommendations if one is aware of the cost-effectiveness of what might be displaced, i.e. existing services.

> *"NICE ought to be spending more time and effort on the other kinds of programmes it looks at. [For example], the public health programme, the clinical guidelines,… a wider range of technologies, existing technologies, things we have been doing for donkey's years and maybe should be doing less of, things we are not doing enough of, like a lot of public health interventions. [The key question should be] what you evaluate, rather than how you evaluate it."*

Those who did not agree that NICE should evaluate the cost-effectiveness of existing services focused on the practical problems with doing so, including the additional resources it would require, as opposed to the principle.

> *"Asking [NICE] to kind of evaluate every drug or device used in the NHS would seem to me to incredibly clog up their ability to deal with new products that are coming on the market."*

Key research directions emerging from interviews

Table 2 reports a list of research directions identified by interviewees as worth pursuing, as well as a selection of quotes to illustrate each idea. These research directions formed the main part of the discussion at the workshop.

Table 2. Key research directions identified by interviewees

Selected research direction	Illustrative interview quote
Examining the clinical threshold[7]	"The approach of trying to look at individual treatment decisions in different therapeutic areas would possibly give you a better idea of where the threshold actually lies, in clinical practice"
Local decision making/priority setting approaches	"It is time we revisited things like PBMA … to study how commissioners make decisions, and how commissioners allocate budgets … How, in particular, they respond to policy initiatives and NICE recommendations and how they decide what gets cut." "Local commissioners do not have great information; they do not have sufficient analytical resources … They are not being as efficient as they could be with the resources that are available. So how do you improve that? … You give people better information."
Examining the cost-effectiveness of existing services	"We should be switching the money from the things that are … above the threshold that we are funding. So that way at least you know that we are improving efficiency even if you are not going to find an optimum"
Improving/expanding the approach by Claxton et al. (2015)	"We can improve on [the] estimates but … it would be an incremental process of getting rather more robust data [and] in a few instances getting more explicitly relevant data" "One would want ideally to look at panel data where you actually have real changes in expenditure and then real health outcomes as a result."
Decision makers' approach	"I would interview … MPs because they are the ones who are represent the values of the [population] … [The value of a QALY] is a question that is way too difficult to ask random people from the population"
Determination of the health budget	"One … high level issue [is that] the … societal value [of a QALY] is almost certainly very much higher than [the] marginal cost … Does that not imply that we do not spend anywhere near enough on healthcare?"

[7] This concept is explained in more detail in *Examining the clinical threshold* (p.11).

Portfolio optimisation as an alternative to assessment against a threshold as part of HTA	"I would certainly say a portfolio approach [is needed at local level], and that is what PBMA is moving towards"
Eliciting the social values of population	"Do we know enough about the factors that enter into the population's [social values]? No, we don't. We have a … hunch about the direction in which they go but we don't really have any easy way of quantifying them and combining that as a [QALY] weight"
Testing of MCDA	"I would like to see multi-criteria methods given same rigorous testing ... ; testing to destruction, almost"
Alternatives to the QALY	"It's worth people going back … to ask that if we were starting again, knowing what we now know about many of the practical difficulties of eliciting people's responses [to the EQ-5D survey] … what kind of measure we would ideally want to have … rather than feeling that we are locked into the QALY as it currently is"

The workshop

The workshop was held on 4 December at Marriot Country Hall, London. There were nine external participants (see Table 1). It was attended by four members of OHE staff (Nancy Devlin, Adrian Towse, Patricia Cubi-Molla and Sarah Karlsberg Schaffer) and an observer from ABPI (Katie Pascoe).

The main objectives of the workshop were to discuss the output of the interviews and to prioritise possible future research that could contribute to developing methodologies and/or empirical evidence to drive forward the policy debate.

The discussion at the workshop can be summarised under the following key headings, which are explored individually below:

- What are we searching for?
- Further exploration of econometric analysis of NHS data
- Informing national level HTA by better understanding the local NHS
- Improving evidence on social values
- Long-term thinking

The transcript was coded to extract quotes relevant to each of these themes.

What are we searching for?

The first key discussion of the workshop was a re-examination of the question, "what are researchers in this area searching for?" There was debate over whether researchers should aim to identify just one threshold, as opposed to a threshold range within which there is the opportunity for deliberation (similar to the current method used by NICE), and which of these is most useful for decision makers to operationalise in HTA.

Another discussion focussed on the issue that a threshold estimated/implemented today is unlikely to be correct in subsequent years, given changes in budgets, health sector inflation, health care productivity and other factors. The implementation of any threshold will lead to decisions that will impact on resource allocation, and NHS performance and productivity, which in itself will alter the threshold. Because the threshold is dynamic, it was suggested that it is necessary to create a process for generating evidence that can be readily updated. This is one of the advantages of the approach taken by Claxton and colleagues: although the methods are complex and there is some missing data, the data that was used for the project is routinely collected by NHS England.

"Part of the trouble is that the threshold is moveable. As the technologies change, as the money changes, as the population changes, as the things that get approved through NICE change, what is affordable changes. We are chasing this ever-receding idea."

In addition, there was broad consensus that there is no one "right" approach to estimating the threshold. Instead, it was suggested that a range of evidence from different angles is required, in order to triangulate across them and obtain a "reasonable" assessment of a value for money threshold. This approach leaves open the question of who would interpret the evidence and decide on the most reasonable threshold or threshold range. The debate included discussion of Appleby et al. (2007)[8], which suggests that there should be an independent "Thresholds Committee", similar to the Monetary Policy Committee, which revisits the evidence at agreed intervals.

Further exploration of econometric analysis of NHS data

The second key theme of the discussion was that it is valuable to continue to explore econometric analysis of NHS data, i.e. extending and improving the Claxton et al. methods. In particular, the discussion focussed on the use of panel data (data on the same PCT/CCG over multiple time periods) to produce a more accurate estimate of the threshold. Note that this is likely to involve mapping/linking of PCT data to CCG data, given the 2013 reconfiguration of the NHS.

It was also suggested that matching techniques could be used to control for differences between PCTs/CCGs and that it might be necessary to use lagged data as part of the analysis. We understand that the authors of the Claxton et al. report have been engaged by the Department of Health to undertake further work to extend the analysis, although there is, at time of writing, no public domain information as to the nature of the work (or whether it involves the use of panel data).

One workshop attendee suggested that data from the Health Survey for England could be used as an indicator for disease morbidity – the Claxton et al. paper makes use only of mortality data and the authors assume that morbidity improves in proportion to the estimated mortality improvement[9].

Moreover, it was the general view of the group that the results of the Claxton et al. paper are interesting from a perspective beyond that of simply attempting to estimate the threshold. In particular, the paper reports vast disparities in the estimated thresholds from one programme budget category to another. The discussion focussed on the desirability of further research to examine the possible drivers of these disparities, including differences in social values, political expediency, inefficiency in the health system and data inaccuracies.[10] It was agreed that understanding these results would require further research and that this research would be valuable for informing HTA decisions at the national level.

"I think [the differences] can only be explained by a combination of three things. One is that the methods of calculation are leading to inaccuracies, but if we understand the nature of those

[8] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1952475/
[9] The proportion comes from the relative shares of morbidity and mortality in separate estimates of disease burden.
[10] The programme budgeting data used for the Claxton et al. work is known to suffer from reliability issues (see Claxton et al., 2015, p.163-164).

inaccuracies we might be able to calculate better. It may be that we identify clear inefficiencies, which … we might, at the local level, be able to use [to improve efficiency] … Or it may be – and I expect it is this – that we learn more about [social] values and that for some services it is not their QALY generation but it is their existence … that is important."

Informing national-level HTA by better understanding the local NHS

More broadly, a key theme of the discussion was that there is the need for research exploring the behaviour and decisions of the NHS at local commissioner and provider level in order to inform national-level HTA. In particular, further research into the political economy of local NHS commissioning would be worthwhile, including developing an improved understanding of what is considered to be value for money and what is taken into account in local commissioning decisions.

There was discussion of a disconnect between economic models of NHS behaviour, where commissioners choose interventions from the "intervention supermarket" according to cost-effectiveness, and what is observed in practice. It was suggested that many local decisions are made on the basis of national/international level imperatives, for example waiting time targets, EU regulations and teaching ratios. The cost per QALY implied by these decisions was not clear. In addition, there is very little cost-effectiveness evidence available to inform local commissioning decisions. This includes disinvestment decisions, where it was suggested that decision makers are likely to prioritise services partially according to political considerations, such as avoiding provoking vocal patient/political stakeholders.

"[Commissioners] are aware of QALYs and … if they had the evidence and wherewithal to … take a different approach, they would, but the fact of the matter is that they have been hit with so many imperatives from the top. It is not just around things like guidance from NICE; … it is waiting times targets; it is their intervention obligations … They just do not have the head space to handle it but, more importantly, even if they wanted to … they have almost no evidence on QALYs for all the stuff that they might want to disinvest from. The evidence that they have to work with is completely imbalanced."

This led to the suggestion that part of the role of research economists working in the field might be to help develop local prioritisation frameworks in order to strengthen commissioning and aid resource allocation decisions, including disinvestment and improving efficiency away from the margin.[11] It was noted that a key challenge would be determining which local decision makers to make contact with, and how exactly to engage with them. It was also suggested that the results of previous PBMA exercises could be used to analyse prioritisation decisions.

Overall, it was agreed that it would be desirable to have greater consistency in local decision making, and greater consistency between local and national decision making. It was suggested that Clinical Commissioning Groups are perceived to be lacking an equivalent rational framework to that currently used at the HTA level.

[11] It should be noted that there have been cases of local decision makers altering services on the basis of evidence on cost-effectiveness, e.g. the reconfiguration of stroke services in London following a review of Health Care in London (the "Darzi" review) and subsequent consultation.

Examining the clinical threshold

The interview process described above identified only one new research direction proposing an alternative method of estimating the cost-effectiveness threshold: analysing marginal individual clinical decisions. This method involves observing the decisions of clinicians in different therapeutic areas in terms of who is being treated and at what level of cost-effectiveness.

At the individual patient level, funding (treatment) decisions are largely made by clinicians. To some degree, they may prioritise their patients according to the level of benefit a treatment provides for that patient, relative to the cost. Therefore, in a particular therapeutic area, one may be able to estimate the cost-effectiveness threshold by identifying the "marginal" patient – the treated patient for whom the treatment is least cost-effective – and the cost-effectiveness of treating them. The extent to which cost-effectiveness influences clinical decision making is potentially a subject for future research.

For example, one could look in general practice at who is being prescribed cholesterol lowering drugs (which may or may not reflect NICE clinical guidelines) or who is being referred for physical therapy or knee/hip replacement surgery, and try to identify the "cut-off points" for each treatment. It was suggested that data on individual treatment decisions could come from the Clinical Practice Research Datalink (CPRD) and that the level of risk for each patient could be estimated using the QRISK calculator.[12]

Considering budget impact

Another key issue that was discussed during the workshop was the role of budget impact considerations in HTA. There was discussion surrounding the new highly cost-effective but expensive treatments for hepatitis C, which have raised short term affordability concerns across the world. NICE does not consider budget impact in its decision making, assuming in effect that decisions are taken at the margin. The issue of how to accommodate new, cost-increasing technologies falls to the payers, i.e. Clinical Commissioning Groups and NHS England. It was suggested that the fact that the budget impact of NICE TAs is considered at the local level but not the national level reinforces the discontinuity between the two sets of decision makers.

In addition, it was highlighted that local responses to the budget impact of NICE recommendations may change over time. It takes time to reorganise services in order to accommodate new guidance and therefore, identifying the opportunity cost of these recommendations is a dynamic problem. An example was given of the NHS's response to the recommendation of new oral anticoagulants, where it was possible to make savings by releasing money from the existing warfarin clinics but, in some cases, such a release of money could not occur fast enough to implement the guidance within three months. There is an issue of rigidity in the system that is not considered at national level HTA.

> *"The basic concept of focusing on net present values in economic evaluation is totally inconsistent with annual budgets against which [NHS managers are] controlled and monitored. An average chief executive [who has overspent] … does not get much credence from saying, 'but in 30 years' time the model shows that there will be a real benefit.'"*

[12] http://www.qrisk.org/

Improving evidence on social values

The NICE process currently attempts to incorporate some assumed societal preferences, most explicitly favouring medicines for patients at the end of their lives. However, a key topic of discussion at the workshop was the extent to which there is a need for further evidence on social values, including age, rarity and severity, and how best they can be reflected in HTA decision making.

A number of participants expressed the view that the methods used to elicit social values, e.g. discrete choice experiments, can lead to unreliable results and are susceptible to framing problems.

> *"Every time you conduct a survey it seems that you get very different results depending on how you … ask the questions."*

It was suggested that researchers could go further to identify which stated preference elicitation methods represent best practice. In particular, ranking methods, as opposed to trade-off methods, were highlighted as a potentially more reliable or consistent source of evidence on social values. In the psychology literature, there is evidence that human beings find it difficult to attach absolute values (e.g. monetary values) to outcomes. The evidence suggests that they are better at attaching *relative* values, and that it is easier to compare outcomes that are similar.

> *"[Survey participants] are trying to give you an answer, they are trying to give you the best answer they can, but they are reaching for all sorts of things to … give them a clue as to the kind of answer they might give that might not be an unreasonable answer, which means that, as you alter the nature of the question, you can phenomenally alter the apparent measure of the output."*

Overall, it was suggested that a mixed-methods approach would be desirable, particularly the concept of validating quantitative results from large sample surveys with the participants of those surveys. In other words, it is valuable to ask participants whether or not they agree with the implications of their own stated preferences.

There was also additional discussion of the merits of performing further research into MCDA methods, which was a "key research direction" highlighted in the interviews. There was no clear consensus among workshop participants regarding the use of MCDA as part of decision making, with some individuals advocating its use in HTA and others expressing their perception that MCDA may be too algorithmic, and would reduce the role of more deliberation.

> *"I do not want to go as far as MCDA but I think I want to go a bit further in the direction of being clear about the factors that are bearing down on a decision."*

The majority of participants agreed, however, that the HTA process would benefit from a more transparent approach to decision making than is used currently. In particular, it was argued that NICE should be clearer with respect to the non-cost-per-QALY evidence it takes into account in its decisions. It was expressed that, from the information that is published, it is often very difficult to understand exactly what factors were taken into account, their respective weights and the degree of uncertainty regarding the evidence.

Moreover, it was agreed that the HTA process already contains elements of MCDA, as the QALY is a measure of the five different components of quality of life used in the EQ-5D. It was also argued that before MCDA can be implemented in a helpful, meaningful way, it is necessary to have consistent, validated evidence on social values (as described above).

Long-term thinking

The final heading section of workshop discussion focuses on research and policy changes that might be important in the long run. First, the group discussed whether it would be possible to "uncouple" drug pricing, HTA and rewards to R&D. Making decisions based on a cost-effectiveness threshold based on the opportunity cost of health gain has the consequence that manufacturers have the incentive to price their medicines up to just below the threshold. This can be seen as sending the correct incentive to reward companies during the patent period, subject to competition. It could also be seen as politicising a technical estimate of the value of health gained were a new technology to be adopted.

Second, the discussion focussed on the high-level socio-political structures in the UK. For example, the observed discrepancy between willingness-to-pay based estimates of the threshold versus shadow price, opportunity cost based estimates indicates that the UK health service budget is too small. It was suggested that economists do not have a great understanding of how the health budget is determined and what causes it to change over time. As the budget constraint is linked directly to the level of the threshold, it would be beneficial to improve the economic evidence surrounding high level budget setting mechanisms to understand, for example, the role of potential health gain from expanding the use of new or existing treatments, or of estimates of productivity gains, in the setting of future budget levels for the NHS.

Discussion and conclusions

Overall, the experts we approached to participate in this study were interested and willing to contribute their time to this topic, and were highly engaged in the discussion and eager to share their ideas. All participants shared the view that the topic is extremely important – and that it is very desirable to improve the evidence base for decisions regarding resource allocation and patients' access to health care.

There was a consensus view, among the workshop participants, that there is no single best or correct approach to empirically estimating the threshold – and that therefore generating complementary evidence from a range of methods would be an appropriate way of informing the selection of the cost effectiveness threshold. Further research is warranted – and this project has identified a number of promising directions which that research might pursue.

Below, we highlight six specific areas for research, based on the output of both the interviews and the workshop.

1. Inferring the threshold from the output of PBMA exercises

It is suggested above that researchers could revisit PBMA and other approaches to local decision making. For example, there may be scope to analyse the output of completed PBMA exercises such as the pilots performed in three sites in Scotland by the team at Glasgow Caledonian University, in partnership with the Scottish Medicines Consortium (SMC), local councils, the Scottish Government and NHS Boards. The pilots involved

examining how resources are currently spent before focusing on marginal benefits and marginal costs of changes in that spend. One could analyse these results to try to understand what is being traded-off by decision makers and the criteria used to make these decisions. While this would be informative, the implications for the national cost effectiveness threshold is not automatic. For example, different priorities may be apparent between different budget holders or between different disease areas.

More broadly, researchers could work with CCGs in England to develop explicit priority setting frameworks. This could help NICE and the NHS in the longer term to have greater coherence/consistency in the way they make decisions that affect resource allocation.

2. Examining the clinical threshold

A second possible area for research is to attempt to identify the threshold by examining treatment decisions made by clinicians. During this discussion, it was recognised that NHS commissioners do not make many of the decisions regarding their budgets but rather act as conduits for resources, signing off bills arising from decisions taken by providers, referrers and patients. Therefore, there is the potential to estimate the threshold by identifying, for a particular treatment or procedure, the point at which a clinician makes the decision to treat or not treat a patient with an intervention.

3. Exploring differences in thresholds across clinical areas

The third specific area for research is an examination of the result found by Claxton et al. that the cost per QALY gained varies hugely across programme budgeting categories. These disparities have multiple potential drivers, including differences in social values, political expediency, inefficiency in the health system and data inaccuracies, which could be explored with further analysis of the programme budgeting data. Understanding these results, through either econometric or qualitative analysis, could help to inform HTA decisions at the national level and could be complementary to an examination of data produced from PBMA exercises, as described above.

4. Further exploration of econometric analysis of NHS data

Closely related to the previous area of research is extending the work completed by Claxton and colleagues. This research could take a number of different routes, some of which are, we understand, already being explored by the authors of the original report.

Specific approaches include the use of panel data (expenditure and health outcome data for the same health units over multiple time periods), which would allow better control of the effects of unobservable differences between health units and the employment of fewer assumptions about their behaviour, and would reflect the lag between health care expenditure and outcomes. This is likely to involve mapping/linking of PCT data to CCG data, given the 2013 reconfiguration of the NHS.

5. Improving evidence on social values

The fifth specific area for research is expanding the evidence base on social values relevant to HTA decision making, such as disease rarity, disease severity, socioeconomic equity and other patient population characteristics. In particular, stated preference studies which have traditionally used trade-off methods could instead (or in addition) make use of ranking or other methods (such as best-worst scaling or other types of discrete choice experiments). These were highlighted as a potentially more reliable or

consistent source of evidence due to their relatively simplicity[13], although this field of research is constantly evolving. It was noted that is it important to ensure these types of experiments are well-funded so that the studies are able to recruit sufficiently large and representative samples.

6. Structured decision making in practice

The final potential research area is to explore the use of structured decision making in practice. Although, there are examples of techniques for structured decision making being used or piloted in HTA in other countries and in other areas of UK health care[14,15], there appears to be a lack of research into how such methods might best be operationalised in the context of NICE Appraisal Committees, how value for money can be incorporated, and the impact on decision making. One option would be for researchers to test the use of existing techniques for structured decision making in mock technology appraisals for hypothetical new products (using current/former Committee members). This type of research is complementary to that aimed at improving evidence on social values.

It should be noted that conducting large-scale public surveys on social values (as described above) is not necessarily required in order use structured decision making methods in HTA. It is also possible to derive the weights in a less resource intensive way, for example, by eliciting the views of the Committee who represent the public and other stakeholders.

The six key research areas identified could help shape the research agenda to optimise reform of the UK HTA decision making framework. Ideally, the final output of this work should contribute to the academic and policy debate around the use of cost effectiveness thresholds in UK HTA processes in future. More broadly, it is desirable to encourage funding bodies, such as NIHR and the MRC Methodology Panel, to prioritise research in these areas.

[13] Craig, B. M., Busschbach, J. J., & Salomon, J. A. (2009). Modelling ranking, time trade-off and visual analogue scale values for EQ-5D health states: A review and comparison of methods. *Medical care*, *47*(6), 634.

[14] Devlin, N. J., & Sussex, J. (2011) Incorporating multiple criteria in HTA. Methods Processes. London: Office of Health Economics

[15] Thokala et al, for the ISPOR MCDA Task Force (2016) MCDA for Health Care Decision Making – An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value in Health (in press)